

УДК 004.738.5+81.32

¹С.С. Федушко, асп.²Ю.О. Сєров, к.т.н., старш. викл.³У.Б. Ярکا, к.т.н., доц.

МЕТОД КОМП'ЮТЕРНО-ЛІНГВІСТИЧНОЇ ВАЛІДАЦІЇ ІНТЕРНЕТ-ІМЕН КОРИСТУВАЧІВ ВЕБ-СПІЛЬНОТИ

Національний університет «Львівська політехніка»

¹E-mail: felomia@gmail.com²E-mail: syerov@ridne.net³E-mail: ul_yarka@yahoo.com

Присвячено актуальній проблемі дослідження та розробки методів валідації інтернет-імен користувачів віртуальних спільнот. Розроблено схему валідації інтернет-імені користувача інтернет-спільноти.

Ключові слова: валідація, віртуальна спільнота, заявник на реєстрацію, інтернет-ім'я.

Постановка проблеми

Розроблення нового підходу до валідації інтернет-імені, яке надає користувач віртуальної спільноти при реєстрації, є актуальним питанням в управлінні системами електронного урядування, дистанційного навчання та модеруванні веб-спільнотами, медіаресурсами, соціальними мережами, енциклопедіями, блогами тощо.

Зі збільшенням чисельності користувачів цих ресурсів у мережі та популярності соціальних комунікацій виникла потреба в розробленні методу перевірки максимальної кількості даних про потенційного учасника веб-спільноти та комп'ютерно-лінгвістичного методу перевірки інформаційного сліду учасника інтернет-спільноти.

Одним із напрямків досліджень є розроблення методів валідації інтернет-імені майбутнього учасника соціальних комунікацій, методу аналізу інтернет-імен у віртуальних спільнотах.

Для спрощення управління та підвищення ефективності діяльності веб-спільнот розроблено основу для комп'ютерно-лінгвістичного методу перевірки інтернет-імені учасника спільноти на відповідність системі правил користувачів цих віртуальних спільнот.

За допомогою комп'ютерно-лінгвістичного аналізу адміністрація спільноти зможе відфільтрувати небажаних учасників та покращити комунікативну поведінку користувачів веб-спільноти.

Особливу увагу приділено проблемі верифікації даних та розробленні методів реєстрації у інтернет-спільнотах, що ґрунтовно впливають на підвищення ефективності функціонування віртуальних спільнот, покращення процедури реєстрації та модерування віртуальної спільноти.

Вирішення цих проблем можливе за допомогою комп'ютерно-лінгвістичного аналізу інтернет-імен учасників веб-спільноти, який здійснюється на основі схеми аналізу інтернет-імен заявників на реєстрацію у віртуальній спільноті.

Збільшення темпів інформаційної діяльності в інтернеті та існуючі на сьогодні проблеми, які виникли на цьому ґрунті, потребують нових підходів до розроблення методів реєстрації, перевірки персональних даних учасників та інформаційного наповнення веб-спільнот.

Особливо важливим є розроблення методів перевірки коректності інтернет-імені користувачів соціальних комунікацій.

Усе це обумовлює актуальність та необхідність розроблення методу комп'ютерно-лінгвістичної валідації інтернет-імен у соціальних комунікаціях.

Аналіз досліджень і публікацій

Аналіз сучасних наукових праць виявив, незважаючи на велику кількість досліджень у цій сфері, вагомий внесок видатних західноєвропейських та американських науковців.

Український сегмент інтернет-простору розвивається доволі швидко. Його інформаційне наповнення потребує дослідження і значного покращення.

В актуальних дослідженнях інформаційної діяльності в соціальних комунікаціях учені розглядають [1]:

- соціолого-психологічні аспекти комунікації людей (K. Popper, A. Podgorecki);
- моніторинг дій користувачів (B. Mobasher, S. Chakrabarti);
- формування спільнот та управління ними (Ю. Серов);
- організацію опрацювання інформаційного наповнення (W. Cunningham, T. O'Reilly);
- позиціонування інформаційного наповнення (И. Ашманов, А. Пелешишин);
- підвищення якості інформаційного наповнення (T. Berners-Lee);
- визначення критеріїв якості інформаційного наповнення (M. Parker).

В Укрнеті ці напрямки досліджень, попри важливість, почали розглядати та аналізувати доволі недавно.

Аналіз досліджень та публікацій [2–8] показав, що кількість фундаментальних досліджень у цій області невелика, проте науковий інтерес зростає і є вже доволі вагомим результатом досліджень.

Мета роботи – розроблення схеми комп'ютерно-лінгвістичної валідації інтернет-імені користувача веб-спільноти.

Для досягнення мети були поставлені такі завдання:

- аналіз існуючих досліджень віртуальних спільнот;
- формування вимог до інтернет-імені учасників спільноти;
- аналіз вимог до інтернет-імен власників існуючих веб-спільнот;

– розроблення методу комп'ютерно-лінгвістичної верифікації інтернет-імен заявників на реєстрацію у веб-спільноті.

Поняття та вимоги до інтернет-імені користувача веб-спільноти

Інтернет-ім'я (нікнейм) для учасника віртуальної спільноти є своєрідною візитною карткою. Від вибору інтернет-імені залежить рівень комунікативної поведінки користувача у спільноті, авторитетність його серед інших учасників спільноти.

На рейтинг спільноти впливає також список учасників спільноти (користувачів, які пройшли реєстрацію у спільноті).

Адміністрація будь-якої віртуальної спільноти в системі правил інтернет-спільноти зазначає вимоги до формату інтернет-імені, яке надає користувач у формі реєстрації веб-спільноти.

Система правил для користувачів інтернет-спільноти [3; 7] – це вимоги (заборони) до контенту спільноти, оформлення повідомлень та облікового запису користувача (у підписах, інтернет-іменах, статусах та на юзербарах, юзерпіках), на основі яких побудовано алгоритм застосування санкцій до учасника веб-спільноти [8].

Таку систему правил для користувачів інтернет-спільноти адміністратори спільноти формують індивідуально, зважаючи на мету створення, сценарій розвитку, аудиторію та тематику віртуальної спільноти.

Оптимальними для ефективного функціонування спільноти є такі вимоги до формулювання інтернет-імені учасників спільноти:

- формат «Ім'я Прізвище» – надання справжнього імені (відповідного паспортним даним) учасником спільноти, формулювання якого збільшує рівень достовірності інформаційного наповнення, зменшує кількість конфліктів, оскільки автор інформаційного наповнення зацікавлений у добрій репутації в інтернет просторі, який є загальнодоступним джерелом інформації;
- нікнейм, який має смисл (у більшості спільнот) – це особисте, переважно, вигадане

ім'я (псевдонім), використовуване користувачем інтернет-мережі в чатах, веб-форумах, на сайтах та вікі-сайтах замість справжнього імені (даного при народженні, зафіксованого в документах) для збереження анонімності в інтернет-комунікації [6];

- заборона використання нецензурної лексики в інтернет-імені;

- довільний вибір інтернет-імені, який є негативним фактором для існування такої спільноти.

Адміністрація не ставить ніяких заборон у виборі інтернет-імені, реєстрація проходить в автоматичному режимі. Така спільнота матиме низький рейтинг та авторитет серед інтернет-користувачів.

Простежуючи тенденції, бачимо, що «віртуальний світ» (анонімність, уявна захищеність користувачів мережі) відходить у минуле, оскільки користувачі переносять свою реальність в інтернет добровільно і невимушено [9].

Отже, для підтримання порядку у спільнотах слід запроваджувати правила вибору та користування інтернет-іменами.

Оптимальними для ефективного функціонування спільноти є вимоги до формулювання інтернет-імені учасників спільноти – формат «Ім'я Прізвище», коли форма реєстрації містить пункти «Ім'я» та «Прізвище», та заборона використання нецензурної лексики в інтернет-імені.

Таким чином, інтернет втрачає повну анонімність, стаючи все більш «реальним», адже коментування, створення дописів виключно під своїм справжнім іменем стає нормою для дедалі більшої кількості користувачів віртуальних спільнот.

Валідація інтернет-імені

Розглянемо детальніше процес перевірки поля інтернет-імені «username», а саме, перевірку на відповідність формату «Ім'я Прізвище» (на основі інтернет-форуму «Форум Рідного Міста»).

Поле «username» – це поле реєстраційної форми користувача віртуальної спільноти, в

яке користувач вказує своє майбутнє інтернет-ім'я в цій спільноті.

У системі правил для користувачів інтернет-спільноти зазначено, що за вимогою адміністрації користувачі повинні при реєстрації подавати власне ім'я та прізвище за форматом «Ім'я Прізвище». Цей пункт системи правил комунікативної поведінки користувачів інтернет-спільноти дає можливість відсіяти небажану аудиторію спільноти, в результаті можливість підтримувати високий рейтинг та рівень достовірності інформаційного наповнення спільноти та сприятливу комунікативну атмосферу серед користувачів спільноти.

У процесі функціонування такої віртуальної спільноти часом утворюється закрыта соціальна спільнота, метою якої є цікава та корисна інтернет-комунікація, обмін інформацією, знаннями, досвідом та здійснення/отримання он-лайн послуг. Важливим критерієм для ідентифікації користувача є його інтернет-ім'я.

Справжнє ім'я (ім'я, яке користувач використовує в реальному житті і відповідне паспортним даним) зобов'язує користувача створювати правдиве та автентичне інформаційне наповнення, створене з дотриманням системи правил користувачів інтернет-спільноти, дає йому можливість користуватися послугами в он-лайн (наприклад, придбати квитки на літак, оплатити комунальні послуги).

Вирішенням проблеми валідації інтернет-імені користувача віртуальної спільноти вбачаємо у створенні методу валідації інтернет-імені учасника веб-спільноти.

Зосередимо увагу на перевірці поля «username» на відповідність формату «Ім'я Прізвище», оскільки вибір такого формату інтернет-імені є найкращим для розвитку, функціонування та модерування веб-спільноти.

В основу перевірки покладено такі бази даних:

- база заборонених слів;
- чорний список заявників на реєстрацію;

– база даних імен учасників веб-спільноти.

Список заборонених слів формується адміністраторами на початку створення веб-спільноти [4; 7; 8].

Прикладом заборонених слів у веб-спільнотах [4] є ненормативна лексика, суржик, лексика, яка не має змістовного навантаження, використання фамільярної лексики, слів, які використовуються в агресивному та зневажливому стилі висловлення думок до інших учасників веб-спільноти тощо.

Чорний список заявників на реєстрацію у веб-спільноті – перелік усіх користувачів спільноти, які вибрали інтернет-ім'я, невідповідне системі правил користувачів веб-спільноти, і при валідації інтернет-імені не забажали змінити дані на коректні. Цей список створений для економії часу адміністраторів, модераторів та зменшення затрат на модерування спільнотою.

База даних інтернет-імен учасників веб-спільноти – список всіх зареєстрованих імен учасників спільноти, що дозволяє автоматично визначити правильність інтернет-імені, мову написання та уникнути дублювання імені учасників веб-спільноти, що не призведе до проблем з ідентифікацією інформаційної діяльності учасників віртуальної спільноти. Адміністратори регулярно доповнюють ці бази даних.

Є три етапи валідації інтернет-імені заявника на реєстрацію у веб-спільноті (див. рисунок):

Перший етап валідації – перевірка інтернет-імені на наявність у чорному реєстраційному списку заявників веб-спільноти. Якщо дані форми реєстрації учасника веб-форуму та інтернет-ім'я не знайдено в цьому списку, то відбувається наступний етап перевірки. Коли ж дані поля «username» вже існують у чорному списку заявників на реєстрацію, то користувачу відмовлено в реєстрації в цьому веб-ресурсі.

Другий етап валідації – аналіз імені користувача форуму форматом «Ім'я Прізвище»

в базі інтернет-імен форуму на наявність імені, яке перевіряємо, для уникнення дублювання імен.

Якщо інтернет-ім'я виявлено в базі імен форуму, то надсилається повідомлення заявнику з проханням змінити інтернет-ім'я або ж з уточненням факту реєстрації раніше. Після зміни даних відбувається перевірка оновленого інтернет-імені заново. Якщо користувач не змінив поданого інтернет-імені, то йому буде відмовлено в реєстрації.

Якщо перевірка відбулась успішно та імені немає в базі імен спільноти, відбувається перевірка імені в базі заборонених слів спільноти.

На третьому етапі валідації у випадку успішної перевірки, тобто заборонених слів не виявлено у формулюванні інтернет-імені заявника, інтернет-ім'я приймається адміністратором для реєстрації і заявнику присвоюється id-учасника, оскільки інтернет-ім'я відповідає системі правил користувачів веб-спільноти.

Якщо заявник на реєстрацію не змінив некоректного інтернет-імені, адміністрація надсилає повідомлення заявнику з вимогою змінити дані. Якщо він ніяк не відреагував на вимогу або ж відмовився змінити дані, то заявника повідомляємо про відмову в реєстрації, попередньо занісши його дані в чорний список заявників на реєстрацію. Якщо користувач змінив дані на вимогу адміністрації, то відбувається перевірка зміненого інтернет-імені заново.

Висновки

Здійснено огляд останніх досліджень у галузі соціальних комунікацій та інформаційної діяльності, який свідчить про актуальність розроблення нових методів валідації даних користувачів віртуальних спільнот.

Запропоновано метод валідації інтернет-імені користувачів віртуальної спільноти та проаналізовано вимоги до інтернет-імен власників веб-спільнот.

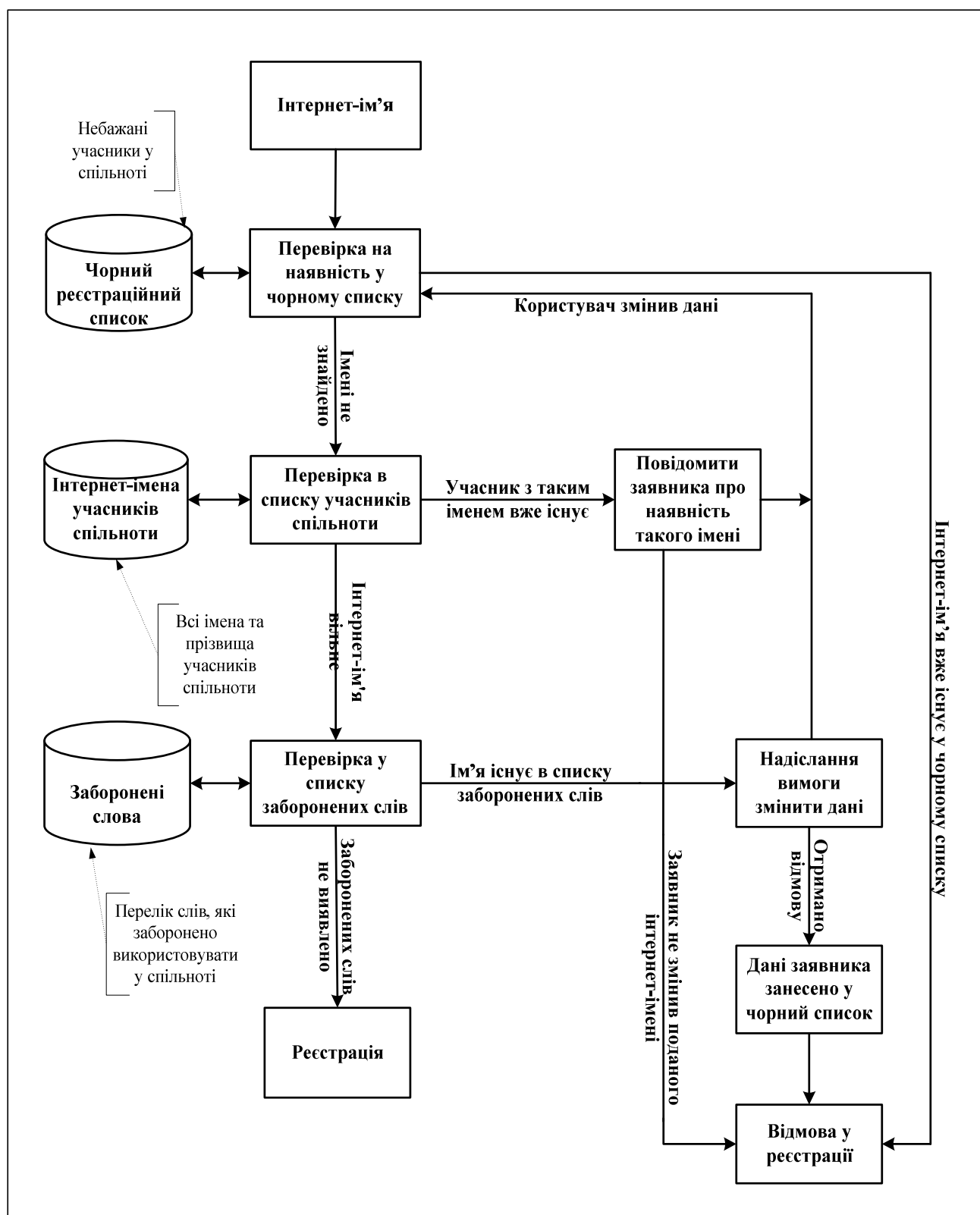


Схема валідації інтернет-імені заявника на реєстрацію у веб-спільноті

Розроблено метод комп'ютерно-лінгвістичної верифікації інтернет-імен заявників на реєстрацію у веб-спільноті на основі схеми валідації інтернет-імені заявника на реєстрацію у веб-спільноті (веб-форумі).

Достовірність персональних даних, фільтрація забороненої лексики та небажаних користувачів стали причинами для розроблення методу комп'ютерно-лінгвістичної верифікації інтернет-імен заявників на реєстрацію у веб-спільноті. Метод комп'ютерно-лінгвістичної валідації значно полегшить роботу модераторам та адміністраторам, зменшить трату часу на модерування ними віртуальної спільноти.

Результати дослідження можуть суттєво допомогти у створенні ефективних методів валідації особистих даних, які надав користувач у своєму обліковому записі.

У майбутніх дослідженнях потрібно розробити алгоритм перевірки всіх полів реєстраційної форми учасника веб-форуму для максимальної достовірності його особистих даних. Особливу увагу приділено комп'ютерно-лінгвістичній перевірці поля інтернет-ім'я «username» у формі реєстрації користувача веб-спільноти на прикладі веб-форуму «Форум Рідного Міста». Цей матеріал є доброю основою для розроблення спеціалізованого програмного забезпечення для перевірки персональних даних реєстраційної форми користувача віртуальної спільноти.

Література

1. *Croll, A.; Power, S.* 2009. Complete Web Monitoring: Watching Your Visitors, Performance, Communities, and Competitors. – O'Reilly Media. – 672 p.

2. *Березко О.* Аналіз сучасних концепцій розвитку середовища WWW / О. Березко, А. Пелешишин // Комп'ютерні науки та інформаційні технології: Вісник НУ «ЛП». – 2006. – № 565. – С. 57–64.

3. *Пелешишин А.* Використання СПКПКІФ для побудови алгоритму застосування санкцій до учасників Веб-спільноти / А. Пелешишин, Ю. Серов, С. Федущко // Conf. on Computer Science and Engineering CSE-2010. – Lviv, 2010. – С. 392–393.

4. *Пелешишин А.* Методи відстеження появи небажаного інформаційного наповнення Веб-форуму / А. Пелешишин, Ю. Серов, С. Федущко // Вісник НУ «ЛП»: ІСМ-Львів: НУ «ЛП», 2010. – № 689. – С. 303–312.

5. *Пелешишин А. М.* Оптимізація форумів та інших форм спільнот користувачів WWW / А. М. Пелешишин // ІСМ: Вісник НУ «ЛП». – 2004. – № 519. – С. 275–284.

6. *Серов Ю.* Використання нечітких множин для моделювання активності учасників Веб-спільнот / Ю. Серов, Р. Кравець // Автоматизированные системы управления и приборы автоматики: всеукраинский межведомственный научно-технический сборник. – Харків, 2007. – № 141. – С. 113–118.

7. *Федущко С.* Комп'ютерно-лінгвістичний метод фільтрації забороненої лексики / С. Федущко, А. Пелешишин, Ю. Серов // Міжнародна науково-технічна конференція «Системний аналіз та інформаційні технології». – К., 2011. – 509 с.

8. *Серов Ю. О.* Методи та засоби побудови ефективних віртуальних спільнот на основі Веб-форумів: автореф. дис. на здобуття наук. ступеня канд. техн. наук: спец. 01.05.03 «Математичне та програмне забезпечення обчислювальних машин і систем» / Ю. О. Серов; НУ «ЛП». – Львів, 2010. – 23 с.

9. *Федущко С.* Аналіз архітектури та сучасних тенденцій розвитку віртуальних спільнот / С. Федущко // Вісник НУ «ЛП»: ІСМ. – Львів: НУ «ЛП», 2011. – № 699. – С. 362–375.